A Place to Go: Locating Damaged Regions after Natural Disasters through Mobile Phone Data

Galo Castillo-López, María-Belén Guaranda, Fabricio Layedra, and Carmen Vaca

Facultad de Ingeniería en Electricidad y Computación, Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Ecuador
{gadacast,mguarand,fblayedr,cvaca}@fiec.espol.edu.ec

Abstract. Large scale natural disasters involve budgetary problems for governments even when local and foreign humanitarian aid is available. Prioritizing investment requires near real time information about the impact of the hazard in different locations. However, such information is not available through sensors or other devices specially in developing countries that do not have such infrastructure. A rich source of information is the data resulting from mobile phones activity that citizens in affected areas start using as soon as it becomes available post-disaster. In this work, we exploit such source of information to conduct different analyses in order to infer the affected zones in the Ecuadorian province of Manabí, after the 2016 earthquake, with epicenter in the same province. We propose a series of features to characterize a geographic area, as granular as a canton, after a natural disaster and label its level of damage using mobile phone data. Our methods result in a classifier based on the K-Nearest Neighbors algorithm to detect affected zones with a 75% of accuracy. We compared our results with official data published two months after the disaster.

Keywords: Mobile Phones Activity · Spatio-Temporal Analysis · Disaster Management

1 Introduction

Natural disasters are unpredictable issues that worldwide governments need to face with a rising trend, each year. The United Nations [25], the World Bank [4, 13] and others [16, 15, 3] have stated the importance of defining more effective ways of disaster management to overcome the consequences of the catastrophe in a time-fashioned way.

Large scale natural disasters involve budgetary problems for governments [5] even when local and foreign cash donations are available. Donations must be invested in a broad geographical area since extreme disasters usually affect a large number of neighboring zones. Prioritizing investment requires near real time information about the impact of the hazard in different locations [26]. Such information is not available through sensors or other devices specially in developing countries that do not have such infrastructure. Moreover, it has been observed that for disasters such as earthquakes, the relation between the country's income and damage is inverse, the less developed the country, the more damage will suffer[17]. In other words, people in developing countries will be more vulnerable. A rich source of information is the data resulting from mobile phone activity that citizens in affected areas start using as soon as they become available after the disaster, hours in some cases. We exploit such source of information in this work to conduct different analyses in order to infer the affected zones after the earthquake that took place in the Ecuadorian province of Manabi on April 16th, 2016. We characterize each of the cantons of Manabi using:

1. Network analysis applied to the towers the individuals who conduct mobile activity are connected to.
2. Temporal analysis to observe how the rythms of activity change after the disaster and;
3. Geographical analysis to propose the *Visitor Diversity Index*, an index based on the different towers involved in the mobile phone events generated by people located in a given area.

Using the aforementioned analyses we propose as our main contribution a classifier based on K-Nearest Neighbors to detect affected zones with a 75% of accuracy. We do so exploiting information of mobile phones activity generated in the following 24 hours after the disaster and the results are compared with data collected through surveys and published officially two months after the disaster.

This article is organized as follows: Section 2 presents previous work regarding disaster management. Section 3 describes the datasets we analyze to build our model. Section 4 describes the metrics we propose as features to characterize each canton in the affected area, as well as the proposed classifier. Section 5 shows the results and discussion. Finally, Section 6 presents the conclusions and future work.

2 Related Work

There are many studies focused on defining new ways to deal with natural disasters, such as: wildfires [21], earthquakes [10, 26, 1, 20], floods [9, 7, 24, 22], typhoons [27], among others [8, 18, 28]. These previous works have proposed a wide variety of solutions going from tools to improve the management of the situation, to data analytics from different sources to extract insights to boost the decision-making process during the disaster.

In the field of disaster management tools, the concept of crowd-sourcing was introduced by the Global Earth Observation Catastrophe Assessment Network (GEO-CAN), formed to facilitate a rapid damage assessment after the 12 January 2010 Haiti earthquake [11]. The tool assisted in quantifying building damage through crowd-sourced imagery with spatial resolutions of up to 50 cm captured by DigitalGlobe, GeoEye, ImageCat and the Rochester Institute of Technology. Another approach was presented by MacEachren et al. [19] with SensePlace2, a map-based web application that creates dynamic geographic, temporal, and thematic visualizations to enhance situational awareness using tweets as source. In the same line, Ashktorab et al. [2] developed Tweedr, a system to extract actionable information for disaster relief workers, analyzing the text of 17 million of tweets collected during 8 years and corresponding to 12 crisis events occurred in North America.

A lot of data sources have been used to address the issues of this research field, specifically the estimation of damage scale and identification of affected areas after the events. In [10], satellite, aircraft and unmanned aerial vehicles(UAVs) data were fusioned to estimate the damage scale of several incidents including the Haiti earthquake of 2010, Hurricane Irene of 2011, Hurricane Sandy of 2012 and the Illinois tornadoes of 2015. Oliveira et al. [21] characterized wildfire-affected areas in Portugal taking as resource: demographic data.

Social media data has also been widely used in this regard. For instance, Cerutti et al. [7] identify disaster affected areas after a flood in Italy in 2013, using geo-spatial footprints from Twitter. Following the same path, Cresci et al. [8] detect mentions of damage among emergency reports on Twitter employing a SVM classifier. Moreover, Yuan and Liu [29] demonstrated that the usage of social media to identify critical affected areas at the county level during disasters is viable. On the other hand, Wilson et al. [26] used Call Detail Records (CDRs) to estimate the displacement of people after the 2015 Nepal earthquake. Pastor-Escuredo et al. [22] calculated the impact of the 2014 France flooding using the same kind of data, just like Andrade et al. [1], with the metric RiSC, that quantifies the infrastructural damage of a region after the 2016 Ecuador earthquake.

Our work aims to continue with the promising results of [8] and [1] but mixing both techniques: a set of classifiers applied to this domain, and the characterization of the affected geographical zones using mobile phones activity.

3 Dataset

In this study, we used data from two different sources and nature, one for our analysis and the other one, as a ground truth.

3.1 Mobile phones activity dataset

Published by a global Telecommunication provider operating also in Ecuador, in the form of CSV files. The purpose of this release was to generate insights regarding the earthquake that stroke Ecuador in April 2016. This dataset is the same used in [1]. These files contain 11 million records, of SMS(Short Message Service) messages and phone calls. The records correspond to two periods:

1. April 15th - 18th: interval that match with one day before and three days after the earthquake.
2. July 15th - 17th: interval that corresponds to 3 months after the catastrophe.

For the purpose of this study, we used records from the first period.

In order to explain what the entries of the dataset represent, the following concepts are needed:

1. Event: either a mobile phone call or a SMS. All the events in the dataset started in any city located in the province of Manabi.
2. Event tower (ET): this is the tower the user's device connected to when generating the *event* .
3. Home tower (HT): this is the tower where the user's device has been connected to most of the times, historically, when generating an *event*.

Events were aggregated by: canton, the nearest *ET's* geographical coordinates, and the hour of the day when the 'event' took place. In particular, an entry portray the volume of *events* produced in the same date, hour, *event tower* and whose transmitters belong to the same *home tower* as shown in Fig. 1.

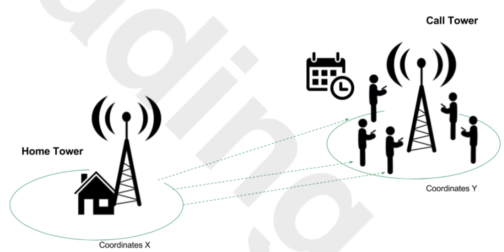

Fig. 1. Event tower where the *event* (either a call or a SMS) takes place and Home tower where the user starting the *event* belongs to as labeled by the Telecommunications provider. Taken from [6].

There are 11 thousand different cell towers, where 808 of them are positioned in the province of Manabi. The towers distribution in Manabi is presented in Fig 2. Note that we do not posses information of the users generating the *events*, since the Telecommunication provider obeys privacy restrictions. Moreover, we do not know if a record corresponds to an SMS or a call.

3.2 Official dataset

The *Secretaría Nacional de Planificación y Desarrollo(SENPLADES)* is an Ecuadorian government's entity in charge of the planning of strategies for the development and well-being of the country. They presented, two months after the disaster, an after-earthquake report including a map which identified the damaged

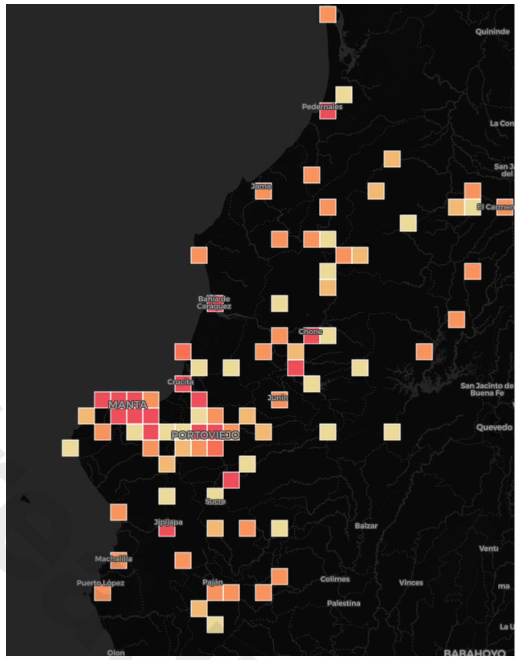

Fig. 2. Event towers distribution in Manabi. The red color represents a major concentration of towers, whereas yellow color represents a minor concentration.

cantons of the province of Manabi, in order to prioritize help to the most affected zones. This map is shown in Fig.3

On the right side, only 3 intensity levels, colored differently, cover the province of Manabi. This is why in [12], a new map is generated, using just 3 levels for simplicity. This new map, as shown in Fig.4, classifies cantons under 3 different levels of affectation. Level 1 (red color) points to the most damaged zones, and corresponds to level 8, yellow matches with level 7 and blue with level 6 in the EMS-98 scale [14].

This map helps us to label cantons according to their damage level. However, for the level 3 class, there are only two elements. Hence, the class 3 is underrepresented, and not useful for any machine learning model. Levels 2 and 3 have a similar affectation description, which is why we decided to merge levels 2 and 3. Thus, we have levels 1 (highly damaged) and 2 (moderately damaged), with 9 and 11 cantons respectively. These labels are used as ground truth for the validation of our proposed model.

4 Methods

The goal of this study is to automatically label a canton or any geographic unit, as highly or moderately damaged after a natural disaster. In order to detect

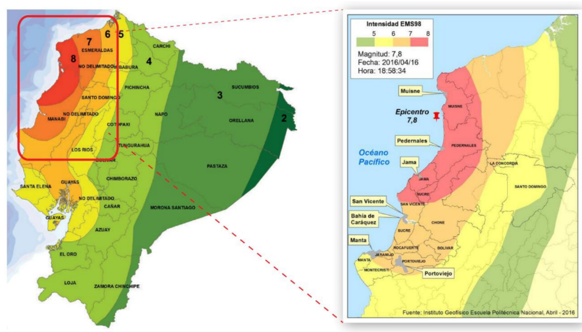

Fig. 3. Affectation intensity levels, corresponding to the province of Manabi, Ecuador. The red color represent an EMS-98 level of 8, while the yellow corresponds to level 6. Results from the Evaluation of the Seism of April 16th, 2016, by SENPLADES.

the level of damage of a given canton, we developed three automated classifiers using records produced by mobile phone data and information that can be quickly calculated such as the distance from a canton to the earthquake epicenter. In this section we describe the methods we used to work with the aggregated mobile phone data described in Section 3. First, we propose three different approaches to characterize the activity presented in each of the Manabí's cantons: spatial, temporal and networks. In doing so, we calculate four metrics that later would be used as features to perform the classification task: visitors diversity index, distance to the earthquake epicenter, time series similarity and eigenvector centrality. Next, we perform three supervised learning algorithms: Linear Support Vector Classification (SVC), Logistic Regression and K-Nearest Neighbors (KNN); using the metrics previously calculated. Finally, we use a Leave-One-Out cross validation in order to evaluate our results.

4.1 Cantons Characterization

Temporal Analysis Aggregating cell phone activity at canton level and by time spans of one hour, where the range of mobile phone events goes from 00h00 to 23h00, could finely represent the daily temporal behavior of the inhabitants of any canton. Whether it exists a significant difference on the activity of a canton between one day before and one day after a natural disaster, then it could be an indicator that it has occurred a disruptive event on that canton. For instance, the canton Manta, which was one of the most devastated cantons, showed an abruptly change in human mobile activity generation, as shown in Fig. 5

Based on that premise, we examine the temporal behavior of Manabí's cantons, with the aim of identifying significant changes between the activity produced on April 15th and April 17th, since the catastrophe took place at the night of April 16th. We aggregate the amount of mobile phone events by canton

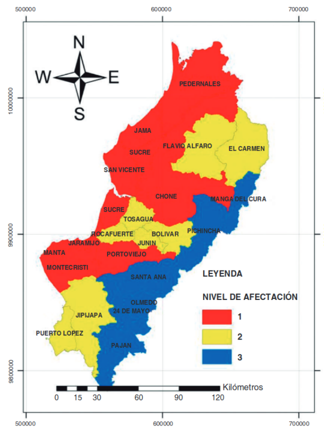

Fig. 4. Affectation intensity levels, corresponding to the province of Manabi, Ecuador. The red color corresponds to the most affected zones, while blue indicates the least affected ones. Taken from [12].

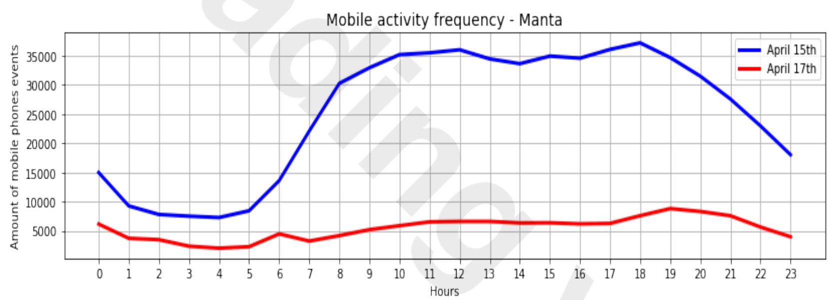

Fig. 5. Mobile activity frequency for April 15th (one day before the earthquake) and April 17th (the day after the catastrophe), of the canton of Manta.

and hour for each day to obtain the daily temporal series. Next, in order to discover how dissimilar the time series of both days are for each canton, we use the Euclidean distance as a metric for calculating the differences of both days on the time domain, as shown on equation 1.

$$dsm = \sqrt{\sum_{i=0}^{23} (a(t_i) - b(t_i))^2} \tag{1}$$

On Equation 1, $a(t_i)$ and $b(t_i)$ represent two points of the two days time series at the same time slot. Through this metric, we can quantify the level of variation of the mobile phone activity for any two days of a particular canton and characterize the ones who present drastic changes in their activity, as well as later identify the causes that could lead to this.

Spatial Analysis Natural disasters cause people to move from one place to another after such events, as mentioned in Section 1. People may move from close or far away places to another because of various reasons: to provide humanitarian aid, searching a safe place to stay, etc. The state-of-the-art has evidenced in a bunch of case studies where mobility patterns have been used to characterize the damage of a region after a natural disaster using cell phone activity. For instance, Andrade et al. [1] concluded that people who travel farther distances during the first 24 hours after an earthquake, come from places that have been less affected than those who travel shorter distances. In a similar way, we propose *Visitors Diversity Index*, a tuned *Shannon Entropy*, which is a metric that explains the popularity of a place P_i in terms of the amount of different places where people, currently located in P_i, come from.

$$VDI(i) = \frac{-\Sigma_{a=1}^{A} \rho_{ia} log(\rho_{ia})}{log(\mathcal{A})} \tag{2}$$

The *Visitors Diversity Index*, as shown in equation 2, is explained as follows: the factor ρ_{ia} is the proportion of an Event Tower i recording mobile phone events from people that have associated the a_{th} Home Tower, for A different Home Towers. A high *Visitors Diversity Index* implies that the area where the Event Tower is located, received people from many different geographic areas. We calculated the Visitors Diversity Index with data produced the day after the earthquake, to identify relevant locations post-disaster. However, the visitors diversity index is calculated for every Event Tower location and we need to characterize locations at canton level. Then we aggregate the indices by canton and calculate the cantonal geometric mean, which indicates the central tendency of a set of visitors diversity values for each canton. Finally, we obtain the *Z score* of the geometric means to normalize the cantonal visitors diversity indices.

Moreover, we also calculate the distance in kilometers between the location of the earthquake epicenter and every canton centroid according to the mobile phone towers locations that belong to each canton. We use the Haversine formula, as shown in equation 3, where $r = 6{,}361$ km is Earth's average radius, ϕ_1 , ϕ_2 and λ_1 , λ_2 are the latitude and longitude of point 1 and 2 in radians, respectively.

$$d = 2\ r\ \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_1+\phi_2}{2}\right) + cos(\phi_1)cos(\phi_2)\sin^2\left(\frac{\lambda_1+\lambda_2}{2}\right)}\right) \tag{3}$$

Networks Analysis Mobile activity is, by nature, represented as a communication network. Graphs that depict communication networks can give us actionable insights about the interactions between the entities involved in the network.

We filter records from April 17th, as we want to capture information from hours after the catastrophe. We build a directed graph $G = (V, E)$, where V is the set of vertexes that represent each Tower (Event or Home Tower) and E is the set of weighted edges, where an edge e_{ij} indicates that a group of k clients from a HT v_i made a call in the CT v_j. We assigned the weight w of the edges following the equation 4:

$$w_{eij} = \frac{log(k) * d}{max(d)} \tag{4}$$

Where d is the distance in kilometers between the nodes v_i,v_j. The weight w_{eij} takes into account the amount of users, as well as the shift in kilometers, for showing an edge as important.It is relevant whether if the shift was large (meaning the users moved large distances from their home) or if the amount of users is large (meaning the place is popular or received affected persons).

The main purpose of our graph analysis is to give a high score to the towers that have the highest influence in the network. For this, we select eigenvector centrality as a metric to extract the most important nodes. This algorithm stands that a node is important if it is linked to other important ones (this is, having a major number of entrant edges). For our case study, an important node represent a tower where mobile activity is generated by displaced people.

4.2 Model Training and Evaluation

To perform the classification task we use and compare the results of three supervised learning algorithms: Linear Support Vector Machine (SVM), Logistic Regression and K-Nearest Neighbors (KNN); using the set of features calculated on Section 4.1. Our aim, is to assign each of the 20 Manabí's cantons to a class: high damage or moderate damage. A k parameter of 5 neighbors presented the best performance when using KNN. All the algorithms we use to perform the classification task uses the four metrics previously calculated: visitors diversity index, distance to the earthquake epicenter, time series similarity and eigenvector centrality.

Due to the limited amount of data points we can use, since we only have information about 20 cantons, we perform a Leave-One-Out cross-validation to evaluate the classification. Leave-one-out cross-validation involves separating the data so that for each iteration we have a single sample for the test data and all the rest forming the training data. In that sense, we predict the damage label of every canton. To measure the classifier performance, we used the following metrics: *Accuracy*, defined in equation 5; *Recall*, as shown in equation 6; and the commonly used *F1-Score* defined in equation 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{7}$$

where,

- **TP** is the number of cantons labeled as highly damaged correctly classified as highly damaged (True Positives).
- **TN** is the number of cantons labeled as moderately damaged correctly classified as moderately damaged (True Negatives).
- **FP** is the number of cantons labeled as moderately damaged erroneously classified as highly damaged (False Positives).
- **FN** is the number of cantons labeled as highly damaged erroneously classified as moderately damaged (False Negatives).

5 Results and Discussion

In this section we present the results of our study and briefly discuss the data, approach and limitations.

Table 1 shows the classification performance for all of our tested classifiers. We note that K-Nearest Neighbors presented the best results in all the performance metrics. Even though we obtained the same accuracy with Linear SVC and Logistic Regression, the SVC's recall metric outperformed the Logistic Regression's sensitivity. In fact, we could only correctly label as highly damaged 3 and 5, out of 9 highly damaged cantons using Logistic Regression and Linear SVC respectively. In contrast, we correctly identify 6 out 9 highly damaged cantons using K-Nearest Neighbors (KNN).

Table 1. Classifiers accuracy, recall and F1 score computations according to equations 5, 6, and 7 respectively.

	Accuracy	Recall	F1 Score
Linear SVC	0.70	0.56	0.63
K-Nearest Neighbors	0.75	0.67	0.71
Logistic Regression	0.70	0.33	0.43

The recall metric, which tells us the percentage of total damaged cantons correctly classified, is of great importance for our framework. We note that KNN is capable of identifying a significant amount of regions that have suffered a high percentage of damage. KNN is a simple classification algorithm, and efficient for a few amount of elements as input. It does not make any assumptions on the data, for which it is useful the non-linear data we use. The poorest performance corresponds to logistic regression, which could mean that the dependent and independent variables relationship is not uniform [23]. The features could have a relationship a linear model cannot support or understand it. A limitation of our study is the amount of data per damage level. The model can be improved if there exists more data that represent each damage level. In addition, the original dataset was already aggregated; the scores values could have been affected by this aggregation, as the information is not so granular or specific to a single entity. Moreover, our ground truth only counts with labels at canton level, then we cannot perform any classification at a finer granularity (e.g. city level).

6 Conclusions and Future Work

In the present research work, we propose a novel model to detect and identify the most damaged zones by an earthquake, at a canton level. Even though other approaches have been focused on human mobility and behavior patterns, they have not dealt with the identification of the levels of affectation of regions. Moreover, there is little research applied to countries in means of development, such as Ecuador. Our approach proposes a machine learning model and four different scores or features that give a useful representation of the cantons and can be obtained after the first 24 hours post-disaster. The model proposed, uses these scores in order to identify the level of damage of the cantons of Ecuador.

The results of this study can be used so that decision-makers at governments can effectively direct the humanitarian aid to places that are truly damaged. We show that our method can correctly label a highly damaged canton as "highly damaged" the 67% percent of times. Also, one of the main contributions of our work is that our method allows to identify highly damaged zones within the first 24 hours after a natural disaster. This is greatly valuable for local governments, which have to prioritize the use of resources to the zones that urgently need them the most. Another contribution is the proposed *Visitors Diversity Index*, which is a tuned version of the popular *Shannon Entropy* metric. This index assigns a high popularity score to a place i in terms of the amount of different places where people, currently located in P_i, come from.

Future work could explore the creation of other metrics and scores, that help to represent the cantons and improve the performances of our methods. Furthermore, the proposed methods could be validated using other datasets corresponding to the same or other natural disaster, so to increase the robustness and inspect how general the model is. For this particular case of study, we could work at a canton level, but in other countries, this high granularity level does not exist. Thus, modifications and enhancements of the proposed methods should be considered. Also, since public institutions such as the Ecuadorian "Secretaría Nacional de Gestión de Riesgos" (Risk Management National Secretary), are interested in having information as the cantons classified by damage during the first eight or twelve hours post-disaster; exploring methods that work on those shorter time spans would be useful in the future. Finally, it is possible to use state-of-the-art algorithms of deep learning to improve the identification of the damage level of each canton.

References

1. Andrade, X., Layedra, F., Vaca, C., Cruz, E.: Risc: Quantifying change after natural disasters to estimate infrastructure damage with mobile phone data. In: 2018 IEEE International Conference on Big Data (Big Data). pp. 3383–3391. IEEE (2018)
2. Ashktorab, Z., Brown, C., Nandi, M., Culotta, A.: Tweedr: Mining twitter to inform disaster response. In: ISCRAM (2014)
3. Bank, C.D.: Cdb, world bank partner to increase disaster resilience through improved procurement (2018), https://www.caribank.org/newsroom/news-and-events/cdb-world-bank-partner-increase-disaster-resilience-through-improved-procurement
4. Bank, T.W.: Disaster risk management (2019), https://www.worldbank.org/en/topic/disasterriskmanagement/overview
5. Cardona, O.D., Ordaz, M.G., Marulanda, M.C., Barbat, A.H.: Estimation of probabilistic seismic losses and the public economic resiliencean approach for a macroeconomic impact evaluation. Journal of Earthquake Engineering **12**(S2), 60–70 (2008)
6. Castillo, G., Layedra, F., Guaranda, M.B., Lara, P., Vaca, C.: The silence of the cantons: Estimating villages socioeconomic status through mobile phones data. In: eDemocracy & eGovernment (ICEDEG), 2018 International Conference on. pp. 172–178. IEEE (2018)
7. Cerutti, V., Fuchs, G., Andrienko, G., Andrienko, N., Ostermann, F.: Identification of disaster-affected areas using exploratory visual analysis of georeferenced tweets: application to a flood event. Association of Geographic Information Laboratories in Europe: Helsinki, Finland p. 5 (2016)
8. Cresci, S., Cimino, A., DellOrletta, F., Tesconi, M.: Crisis mapping during natural disasters via text analysis of social media messages. In: International Conference on Web Information Systems Engineering. pp. 250–258. Springer (2015)
9. De Albuquerque, J.P., Herfort, B., Brenning, A., Zipf, A.: A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. International Journal of Geographical Information Science **29**(4), 667–689 (2015)
10. Fernandez Galarreta, J., Kerle, N., Gerke, M.: Uav-based urban structural damage assessment using object-based image analysis and semantic reasoning. Natural hazards and earth system sciences **15**(6), 1087–1101 (2015)
11. Ghosh, S., Huyck, C.K., Greene, M., Gill, S.P., Bevington, J., Svekla, W., DesRoches, R., Eguchi, R.T.: Crowdsourcing for rapid damage assessment: The global earth observation catastrophe assessment network (geo-can). Earthquake Spectra **27**(S1), S179–S198 (2011)
12. Gil, H.A.P.: Efectos del sismo del 16 de abril de 2016 en el sector productivo agropecuario de manabí. La Técnica (17), 30–42 (2017)
13. Giugale, M.: Time to insure developing countries against natural disasters (2017), https://www.worldbank.org/en/news/opinion/2017/10/11/time-to-insure-developing-countries-against-natural-disasters
14. Grünthal, G.: European macroseismic scale 1998. Tech. rep., European Seismological Commission (ESC) (1998)
15. Guha-Sapir, D., Hargitt, D., Hoyois, P.: Thirty years of natural disasters 1974-2003: The numbers. Presses univ. de Louvain (2004)
16. Hoeppe, P.: Trends in weather related disasters–consequences for insurers and society. Weather and climate extremes **11**, 70–79 (2016)

17. Kellenberg, D., Mobarak, A.M.: The economics of natural disasters. Annu. Rev. Resour. Econ. **3**(1), 297–312 (2011)
18. Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., Cebrian, M.: Rapid assessment of disaster damage using social media activity. Science advances **2**(3), e1500779 (2016)
19. MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J.: Senseplace2: Geotwitter analytics support for situational awareness. In: 2011 IEEE conference on visual analytics science and technology (VAST). pp. 181–190. IEEE (2011)
20. Olen, S., Bookhagen, B.: Mapping damage-affected areas after natural hazard events using sentinel-1 coherence time series. Remote Sensing **10**(8), 1272 (2018)
21. Oliveira, S., Zêzere, J.L., Queirós, M., Pereira, J.M.: Assessing the social context of wildfire-affected areas. the case of mainland portugal. Applied geography **88**, 104–117 (2017)
22. Pastor-Escuredo, D., Torres, Y., Martinez, M., Zufiria, P.J.: Floods impact dynamics quantified from big data sources. arXiv preprint arXiv:1804.09129 (2018)
23. Ranganathan, P., Pramesh, C., Aggarwal, R.: Common pitfalls in statistical analysis: Logistic regression. Perspectives in clinical research **8**(3), 148 (2017)
24. Rosser, J.F., Leibovici, D., Jackson, M.: Rapid flood inundation mapping using social media, remote sensing and topographic data. Natural Hazards **87**(1), 103–120 (2017)
25. Schlein, L.: Un: Most deaths from natural disasters occur in poor countries (2016), https://www.voanews.com/a/un-says-most-deaths-from-natural-disasters-occur-in-poor-countries/3548871.html
26. Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., et al.: Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 nepal earthquake. PLoS currents **8** (2016)
27. Yabe, T., Sekimoto, Y., Sudo, A., Tsubouchi, K.: Predicting delay of commuting activities following frequently occurring disasters using location data from smartphones. Journal of Disaster Research **12**(2), 287–295 (2017)
28. Yu, M., Yang, C., Li, Y.: Big data in natural disaster management: a review. Geosciences **8**(5), 165 (2018)
29. Yuan, F., Liu, R.: Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane matthew case study. International journal of disaster risk reduction **28**, 758–767 (2018)